

CAS Natural Language Processing

Study Plan

15.09.2022 The CAS Natural Language Processing is a university study program leading to a “Certificate of Advanced Studies in Natural Language Processing” awarded by the University of Bern as laid out in the regulations for the Studiengänge in Extended Intelligence of 2021-12-09.

1. Course objectives

About The field Natural Language Processing (NLP) has exploded in recent years. It belongs to both computational linguistics as its engineering domain and artificial intelligence as an increasingly important subdomain. In particular the applications based on deep neural networks have reached a performance level which cannot be ignored by any field that is processing natural languages. Machine translation, information extraction, question answering and non-sense natural language generation performed by NLP algorithms are already part of everyday life. However, the new semantic understanding capabilities seen in recent trained models not only expand the NLP application space enormously, but also represent possible first steps towards consciousness and mind outside human and animal nerve systems.

The program targets practitioners who aim for an overview of the NLP domain with main focus on recent developments (deep learning models) and hands-on learning. Some programming experience or the willingness to learn basic scripting is a prerequisite.

The CAS format is designed to align with the participants’ main professional and study activities. The workload is on average about 20% of a full time position over a year. The teaching and learning approaches are team and discussion oriented, aimed at developing practical competency. All modules can also be attended online.

Competence objectives The course competencies are developed in six modules. At the end graduates will:

1. have an overview of the NLP domain and common applications,
2. be able to perform relevant preprocessing tasks needed for advanced NLP,
3. be able to understand neural networks and practice them on own NLP applications,
4. be able to understand transformers and practice transfer learning with transformers for own applications,

5. know discussions related to philosophical and ethical aspects around NLP and artificial intelligence
6. be familiar with active research in the NLP domain.

2. Duration, outcomes and objectives of the modules

Duration and scope

The CAS consists of about 18 days of courses (126 hours attendance) and 16 ECTS credit points (total effort about 480 hours) whereof 4 ECTS result from the CAS project work.

The program consists of four modules in weekly blocks and two modules on eight to ten half days. All modules support distant learning.

Attendance at one introductory event *About the CAS Natural Language Processing* before course admission is mandatory.

Module 1

NLP Fundamentals

In which linguistic and machine learning concepts are introduced, and an overview of the NLP field and common applications is given.

ECTS	2 ECTS credit points (incl. self-studies and project)	Duration	4.5 days + essay writing
Assessment	Written essay	Req. attendance	80%
Learning outcomes	Graduates will <ul style="list-style-type: none"> ● have an understanding of the field of linguistics and study of languages ● understand different tasks of NLP and their requirements ● know most common NLP frameworks and tools ● know the history of NLP in computer science and the humanities ● develop a first project involving NLP methods, which can be used as foundation for the CAS project work ● be able to reflect upon the intertwining of NLP with machine learning and artificial intelligence 		
Learning objectives	<ul style="list-style-type: none"> ● Basic linguistics ● Common NLP tasks ● NLP frameworks ● NLP history ● NLP concept development 		
Learning and teaching methods	<ul style="list-style-type: none"> ● Onsite classes with online transmission ● Lectures, self studies, team work, discussions ● Essay assessment 		
Prerequisites	Own laptop		
Language	English		

NLP Preprocessing and Basic Analysis

In which we learn to perform basic preprocessing and analysis of natural languages.

ECTS	2 ECTS credit points (incl. self-studies and project)	Duration	3 days + project
Assessment	Project presentation	Req. attendance	80%
Learning outcomes	Graduates will/can <ul style="list-style-type: none"> perform basic preprocessing and segmentation of text for NLP purposes perform basic information extraction (know forms of annotation and corresponding evaluation) perform basic critical text analysis (epistemological foundations and theoretical classification) 		
Learning objectives	Basic preprocessing techniques for text data: <ul style="list-style-type: none"> cleaning special characters and numbers removing punctuation, misspells and contractions etc Sequence creation: <ul style="list-style-type: none"> using tokenizers padding sequences embedding enrichment Information extraction and critical text analysis <ul style="list-style-type: none"> Named Entity Recognition Part-of -Speech Tagging 		
Learning and teaching methods	<ul style="list-style-type: none"> Onsite classes with online transmission Lectures and inverted classroom with hands-on tutorials Teamwork on projects Project presentations with peer feedback 		
Prerequisites	Previous modules		
Language	English		

NLP Neural Networks

In which we study and learn how neural networks work, are trained, tuned, assessed and applied for NLP tasks.

ECTS	2 ECTS credit points (incl. self-studies and project)	Duration	3 days = 21 hrs attendance
Assessment	Project presentation	Required attendance	80%
Learning outcomes	Graduates will/can <ul style="list-style-type: none"> know different neural network architectures understand the role of the various network components train neural networks evaluate the performance of neural network models apply neural networks to NLP tasks understand the principles of recurrent neural networks 		
Learning objectives	<ul style="list-style-type: none"> Neural network architectures and components for NLP Model training 		

	<ul style="list-style-type: none"> ● Model predictions and analysis ● Model performance ● Transfer learning ● Model application to sequence-to-sequence tasks, text analytics, text comprehension ● Long-term dependencies
Learning and teaching methods	<ul style="list-style-type: none"> ● Onsite classes with online transmission ● Lectures and inverted classroom with hands-on tutorials ● Teamwork on project ● Project presentations with peer feedback
Prerequisites	Previous modules
Language	English

Module 4

NLP Transformers

In which we study transformers and learn why they have changed the NLP field.

ECTS	2 ECTS credit points (incl. self-studies and project)	Duration	3 days = 21 hrs attendance
Assessment	Project presentation	Req. attendance	80%
Learning outcomes	Graduates will/can <ul style="list-style-type: none"> ● understand the inner workings of the transformer architecture including the attention mechanism and positional embeddings ● understand the differences between autoregressive and autoencoding language models ● know current limitations of transformer-based models such as the input length ● apply a modern transformer-based pipeline to various NLP tasks 		
Learning objectives	<ul style="list-style-type: none"> ● Attention mechanism ● Positional embeddings ● Autoregressive and Autoencoding Language Models ● BERT and GPT as examples ● Encoders and Decoders ● Current limitations 		
Learning and teaching methods	<ul style="list-style-type: none"> ● Onsite classes with online transmission ● Lectures and inverted classroom with hands-on tutorials ● Teamwork on project ● Project presentations with peer feedback 		
Prerequisites	Previous modules		
Language	English		

Module 5

NLP Philosophical and Ethical Aspects

In which we study and discuss ethical and philosophical aspects related to machines being capable of natural language processing

ECTS	2 ECTS credit points (incl. self-studies)	Duration	3 days = 21 hrs attendance
Assessment	Written essay	Required attendance	80%
Learning outcomes	Graduates will/can <ul style="list-style-type: none"> • have an overview of the history and the philosophy of artificial intelligence • know philosophical and scientific presuppositions of artificial intelligence • can relate techniques of AI to well-known scientific methods • know main philosophical discussions on artificial intelligence • know main moral challenges related to artificial intelligence and can discuss solutions from the perspective of ethics • master best practices for ethics dealing with artificial intelligence 		
Learning objectives	<ul style="list-style-type: none"> • Philosophical conceptions of AI, weak vs. strong AI • Extended mind hypothesis • Philosophical concepts of data and data analysis • AI and scientific inference • Ethical challenges due to AI in the light of ethical theories • Machine ethics 		
Learning and teaching methods	<ul style="list-style-type: none"> • Onsite classes with online transmission • Lectures, discussions and independent studies • Essay writing 		
Prerequisites	Module 1		
Language	English		

Module 6

NLP Frontier and Applications

In which we use knowledge from the previous modules to focus on prominent NLP topics and applications with a feasibility study for our final CAS project.

ECTS	2 ECTS credit points (incl. self-studies and project)	Duration	3 days = 21 hrs attendance
Assessment	Project presentation	Req. attendance	80%
Learning outcomes	Module graduates <ul style="list-style-type: none"> • can design, formulate and communicate an NLP project • know current research topics within the NLP field • apply newer NLP approaches to own text 		
Learning objectives	In addition to the final CAS project planning, objectives may be among others <ul style="list-style-type: none"> • Information retrieval • Search engines • Multilingual NLP • Domain-Specific NLP, e.g. legal NLP 		

Learning and teaching methods	<ul style="list-style-type: none"> ● Onsite classes with online transmission ● Lectures and inverted classroom with hands-on tutorials ● Teamwork on project ● Project presentations with peer feedback
Prerequisites	Previous modules
Language	English

CAS Project Work

CAS Project Work

In which participant teams perform and present a comprehensive Natural Language Processing project based on all six CAS modules.

ECTS	4 ECTS credit points	Duration	120 hours
Assessment	Exhibition work	Req. attendance	None
Learning outcomes	Graduates will be able to perform and present a complex Natural Language Processing project.		
Learning objectives	Selection of objectives from all modules. Furthermore, graduates will create an enticing presentation that reports on the practical work carried out, combined with comparison to similar state-of-the-art applications/proof-of-concepts.		
Teaching methods	Project work in teams with remote supervision		
Prerequisites	Module 1 - 6		
Language	English		

Assessment

3. Performance assessment

The performance assessment includes 80% attendance and passed performance assessments from all modules and the CAS project.

- For Module 1 a written essay is assessed.
- For Module 2, 3, 4 and 6 project presentations are assessed.
- For Module 5 a written essay is assessed.
- For the CAS project a report, publication or an app with a presentation is assessed.

4. Final regulations

Entry into force The present plan shall enter into force on 01.10.2022.

15.09.2022 Released by the program management:

Prof. Dr. Thomas Wihler

06.10.2022

Adopted by the Faculty of Science, University of Bern

The Dean

Prof. Dr. Marco Herwegh